10/521881

Improved Inter-Processor Communication System for Communication between Processors

5

Field of the invention

The present invention concerns generally the communication between two or more processors. In particular, the present invention concerns the inter-processor communication between processors that are arranged on the same semiconductor die.

10   Background of the invention

As the demand for more powerful computing devices increases, more and more systems are offered that comprise more than just one processor.

For the purposes of the present invention, a distinction is to be made between computer systems that comprise two or more discrete processors and systems where two or more

15   processors are integrated on the same chip. A computer with a main central processing unit (CPU) on a mother board and an algorithmic processor on a graphics card is an example for a computer system with two discrete processors. Another example of a computer system with several discrete processors is a parallel computer where an array of processors is arranged such that an improved performance is achieved. For sake of simplicity, systems on a board with two or

20   more discrete processors are also considered to belong to the same category.

There are systems where two or more processors are integrated on the same chip or semiconductor die. A typical example is a SmartCard (also referred to as integrated circuit card) that has a main processor and a crypto-processor on the same semiconductor die.

As small handheld devices are becoming more and more popular, the demand

25   for powerful and flexible chips is increasing. A typical example is the cellular phone which in the beginning of its dissemination was just a telephone for voice transmission (analogue communication). Over the years additional features have been added and most of today's cellular phones are designed for voice and data services. Additional differentiators are wireless application protocol (WAP) support, short message system (SMS), and multimedia message

30   service (MMS) functionality, just to name some of the more recent developments. All these features require more powerful processors and quite often even dual-processor or multi-processor chips.

In the future, systems handling digital video streams for example will become

available. These systems also require powerful and flexible chip sets.

Other examples are integrated circuit cards, such as multi-purpose JavaCards, small handheld devices, such as palm top computers or personal digital assistants (PDAs), video and audio devices, devices for use in automotives, and so forth.

5   It is essential for such dual-processor or multi-processor chips that there exists a communication channel for efficient inter-processor communication. The expression "inter-processor communication" is herein used as a synonym for any communication between a first processor and/or system resources associated with this first processor and a second processor and/or system resources associated with this second processor. A shared memory (e.g., a random
10   access memory) is an example of a system resource that usually needs to be accessible by all processors of a chip.

System resources have to be shared in an efficient manner in dual-processor or multi-processor chips where the processors operate in parallel on the same aspect of a task or on different aspects of the same task. The sharing of resources may also be necessary in applications
15   where processors are called upon to process related data.

An example of a multi-processor system is given in the European Patent application EP 0 580 961-A1, filed on 16 April 1993. This Patent application concerns a system with multiple discrete processors and a global bus that is shared by all these processors. Enhanced processor interfaces are provided for linking the processors to the common bus. Such
20   multi-processor systems with a global bus cannot be realized using RISC processors, due to the high bus load which would have an impact on the system's performance. The multi-processor system presented in EP 0 580 961-A1 is powerful but complicated and expensive to implement. The shown structure cannot be used in multi-processor systems on a common die.

Another system is proposed in US Patent US 4,866,597, filed on 26 April 1985.
25   This US Patent concerns a multi-processor system where each processor has its own processor bus. Data are exchanged between these processors via first-in-first-out data buffers (FIFO) which directly interconnect the respective processor buses. It is a disadvantage of this approach that the size of the buffers increases dramatically with the amount of data to be transferred.

US Patent 5,093,780 concerns an inter-processor transmission system that has a
30   data link which automatically reads and writes transfer data. A direct memory access (DMA) unit and a transmitter are assigned to a first processor and a receiver together with a DMA unit are assigned at a second processor. The processor has to set up the transfer by programming the corresponding DMA. That is, the processor has to know upfront whether data are to be transferred. This is a disadvantage of the described inter-processor transmission system, since the
35   respective processor needs to be involved. Another disadvantage of the said system is the fact

that the whole transmission is mono-directional, i.e.,, the implementation is asymmetric. It is just possible to transfer data from the memory 16 on the left hand side of Figure 4 to the memory 26 on the right hand side.

5   A DMA controller for a multi-microcomputer system is disclosed in US patent 5,222,227. The DMA controller has the function of controlling data transfer operations that are executed by the microcomputer systems. Separate address and data pipelines are provided. Tri-State-Technology is used for the buses. The buses CDB and SDB are at least temporarily electrically interconnected. As a consequence, both buses have to be operated at the same clock speed and both buses have to have the same bus width. According to the US patent 5,222,227,

10   only homogeneous buses can be interconnected. There is no external DMA channel used in the system presented.

A multi-processor system with a shared memory is described and claimed in US Patent US 5,283,903, filed on 17 September 1991. The system in accordance with this US Patent comprises a plurality of processors, a shared memory (main memory), and a priority selector

15   unit. The priority selector unit arbitrates between those processors the request access to the shared memory. This is necessary, since the shared memory is a single-port memory (e.g., a random access memory) that cannot handle simultaneous and competing requests from several processors. It is a disadvantage of this approach that the shared memory is expensive as only intermediate storage. The shared memory can get large with high data transfer.

20   Another multi-processor system is described in US Patent US 5,289,588, filed on 24 April 1990. The processors are coupled by a common bus. They can access a shared memory via this common bus. A cache is associated with each processor and an arbitration scheme is employed to control the access to the shared memory. It is a disadvantage of this approach that the cache memory is expensive as only big caches give a real performance boost.

25   In addition, bus conflicts lead to a reduced performance of each processor.

A microprocessor architecture is described in the PCT Patent application PCT/JP92/00869, filed on 7 July 1992, and published under PCT Publication number WO 93/01553. The architecture supports multiple heterogeneous processors which are coupled by data, address, and control signal buses. Access to a memory is controlled by arbitration circuits.

30   Some of the known multi-processor systems use architectures where the inter-processor communication occupies part of the processor's processing cycles. It is desirable to avoid this overhead and to free-up the processor's processing power in order to be able to better exploit the processor's capabilities and performance.

Other known schemes cannot be used for integrated multi-processor systems

35   where two or more processors are located within the same chip.

It is yet another disadvantage of some known systems that they are asymmetric in their implementation which means that different implementations are required for each processor. Furthermore, the effort for formal verification is greater for asymmetric than for symmetric implementations.

5

Summary of the Invention

It is an object of the present invention to provide a scheme for efficient data transfer between two or more processors and/or their associated components.

It is an object of the present invention to provide an inter-processor data transfer
5   scheme that is suited for the integration into a semiconductor die.

These and other objectives are achieved by the present invention which provides a system that comprises at least two integrated processors. According to the present invention, these two processors are operably connected via a communication channel for exchanging information. One processor (P1) has a processor bus, a shareable unit, and a DMA unit with two
10 external DMA channels. The DMA unit and the shareable unit are connected to the processor bus. The other processor also has a shareable unit and a DMA unit with two external DMA channels. Programmable units are employed enabling the processor to set-up the desired communication links. Due to this arrangement, two bi-directional communication channel are establishable between the two bus regimes.

15   The two or more processor can be arranged on a common semiconductor die. This allows to realise computing devices, such as PDAs, handheld computers, palm top computers, cellular phones, and cordless phones, for example.

The communication channel can be used advantageously for communication between two or more processors and/or their associated components. The inventive arrangement
20 suits general multi-core communication needs. The arrangement is highly symmetrical and it allows to minimise the number of otherwise needed bus masters for each processor. The present scheme is expandable and very flexible.

These and other aspects of the invention will be apparent from and elucidated with reference to the embodiment(s) described hereinafter.

Brief description of the drawings

For a more complete description of the present invention and for further objects and advantages thereof, reference is made to the following description, taken in conjunction with the accompanying drawings, in which:

5   FIG. 1 is a schematic block diagram of a dual-processor computer system, according to a

first embodiment of the present invention.

FIG. 2 is a schematic illustration of an inter-processor communication system according to the present invention.

10   FIG. 3 is a detailed block diagram of the inter-processor communication system of Figure 2.

FIG. 4 is a schematic block diagram of a dual-processor computer system, according to

another embodiment of the present invention.

15   FIG. 5 is a schematic block diagram of a dual-processor computer system, according to

another embodiment of the present invention.

FIG. 6 is a detailed block diagram of the DTU unit of Figure 5.

DESCRIPTION OF PREFERRED EMBODIMENTS

The present invention is described in connection with several embodiments.

As shown in Figure 1, a dual-processor system to which the present invention is applied comprises a first processor P1 that is connected via a first processor bus 10 to a first
5   shareable unit 13. A processor bus (also called microprocessor bus) is the main path connecting to the computer system's processor. An example of a shareable unit 13 or 23 is a shared memory (e.g., a random access memory; RAM). The first processor bus 10 is a 64 bit, 20MHz bus. The system comprises a second processor P2 that also has a processor bus 20. This second processor bus 20 is a 64 bit, 66MHz bus. An interconnection between the two processor environments 18
10   and 28 (schematically illustrated by ovals in Figure 1) is established via two bi-directional communication channels 11 and 21. The first bi-directional channel 11 is programmable by the processor P1, as indicated by the arrow 12, and the second channel 21 is programmable by the processor P2, as indicated by the arrow 22. The two bi-directional channels 11 and 21 are hereinafter referred to as intercore communication system 9.

15   More details of the first embodiment are depicted in Figure 2. The intercore communication system 29 comprises a first DMA unit 45 (DMA1) with a first and a second external DMA channel 46, 47. The first DMA unit 45 is connectable to the first processor bus 10 via an internal DMA channel 49. It furthermore comprises a first double tandem unit (DTU) 34 (DTU1) which is connectable via the first external DMA channel 47 to the first DMA unit 45.
20   The DTU unit 34 is programmable by the first processor P1, as indicated by the arrow 32, and the first DMA unit 45 is programmable by the first processor P1, as indicated by the arrow 132 In addition, the intercore communication system 29 comprises a second DMA unit 35 (DMA2) and a second DTU unit 44 (DTU2). The DMA unit 35 has a first and a second external DMA channel 36, 37, and an internal DMA channel 39. The second DMA unit 35 is connected via the
25   internal DMA channel 39 to the second processor bus 20. The second DMA unit 35 and the second DTU unit 44 are connectable via the first external DMA channel 37. The DTU unit 44 is programmable by the second processor P2, as indicated by the arrow 42, and the second DMA unit 35 is programmable by the second processor P2, as indicated by the arrow 142 A first bi-directional communication channel is implemented by the first DTU 34 and a second bi-
30   directional channel is implemented by the second DTU 44. Each DTU 34, 44 is directly connectable to the processor for programming/configuration purposes, as illustrated in Figure 2. An interconnection between the two processor environments 38 and 48 (schematically illustrated by ovals in Figure 2) is established by a bi-directional data transfer.

It is a novel feature of the embodiment given in Figures 1 and 2 that the
35   programming of one DTU unit, e.g. DTU1, allows (bi-directional) data transfer from the

processor environment 48 to the processor environment 38 and vice versa without the programming of any other resource. Data can be moved to the DMA1 and fetched from the shareable unit 23 that is attached to the second processor bus 20. The DTU2 is able to move data to the DMA2 and to fetch data from the shareable unit 13 that is attached to the first processor
5   bus 10.

The dual-processor arrangement illustrated in Figures 1 or 2 allows the second processor P2 to access the shareable unit 13. The shareable unit 23 is accessible by the processor P1.

In more general terms, one processor (processor P2 in the present embodiment)
10   of a multi-processor system in accordance with the present invention is able to access resources (the shareable unit 13 in the present embodiment) that are associated with another processor (processor P1 in the present embodiment). A resource of another processor on a remote bus may be accessed for data up and download from cheap remote memory, for instance. A processor may for example access the memory of a co-processor to fetch data that were computed by the co-
15   processor. These are just two typical examples of situations where a first processor accesses resources on a remote bus.

Various types of processors can be interconnected using the present scheme. It allows to realise chips with multiple homogeneous processors or even with multiple heterogeneous processors. The word processor is herein used as a synonym for any processing
20   unit that can be integrated into a semiconductor chip and that actually executes instructions and works with data.

Complex instruction set computing (CISC) is one of the two main types of processor designs in use today. It is slowly losing popularity to reduced instruction set computing (RISC) designs. The most popular current CISC processor is the x86, but there are also 68xx,
25   65xx, and Z80s in use.

Currently, the fastest processors are RISC-based. There are several popular RISC processors, including Alphas (developed by Digital and currently produced by Digital/Compaq and Samsung), ARMs (developed by Advanced RISC Machines, currently owned by Intel, and currently produced by both the above and Digital/Compaq), PA-RISCs
30   (developed by Hewlett-Packard), PowerPCs (developed in a collaborative effort between IBM, Apple, and Motorola), and SPARCs (developed by Sun; the SPARC design is currently produced by many different companies).

ARMs are different from most other processors in that they were not designed to maximise performance but rather to maximise performance per power consumed. Thus ARMs
35   find most of their use on hand-held machines and PDAs. .

In the above sections some examples of the processors were given that can be interconnected in accordance with the present invention. Also suited are Digital Signal processors (DSPs), the processor cores of any of the known processors, and customer specific processor designs. In other words, the present concept is applicable to most microprocessor

5   architectures. One can even interconnect a processor with a slow processor bus and a processor with a fast processor bus.

For the purpose of the present application, the following is also considered to be a processor: central processing unit (CPU), microprocessor, digital signal processor (DSP), system controller (SC), co-processor, auxiliary processor, control unit and so forth.

10   A direct memory access (DMA) unit is a unit that is designed for passing data from a memory to another device without passing it through the processor. A DMA typically has one or more dedicated internal DMA channels and one or more dedicated external DMA channels for external peripherals. Such an external DMA channel – contrary to an internal DMA channel that is controlled by the processor to which it is associated – is set-up by external agents

15   in order for the remote processor to get access to another processor's shareable unit. For instance, a DMA allows devices on a processor bus to access memory without requiring intervention by the processor.

Examples of shareable units are: volatile memory, non-volatile memory, peripherals, interfaces, input devices, output devices, and so forth.

20   The intercore communication system, according to the present invention, decouples the data flow between the clock domain of a first processor P1 and the clock domain of a second processor P2. This means that within the limits of the inventive data transfer system, the activity on one processor does not require simultaneous and equivalent activity on the other processor.

25   Details of the intercore communication system 29, according to the present invention, are described in connection with Figure 3. The intercore communication system 29 comprises a first DTU 34 (DTU1), a second DTU 44, (DTU2), a first DMA 45 (DMA1), and a second DMA 35 (DMA2). In the present embodiment, the DMA unit 35 comprises two external DMA channel units 56, 57. The internal channel 39 of these two external DMA channel units

30   56, 57 is connected to the processor bus 20.

The first external DMA channel unit 56 is connected via a link 36 to the second DTU 44. The second external DMA channel unit 57 is connected via a link 37 to the first DTU 34. The first DMA unit 45 comprises two external DMA channel units 54, 55. The internal channel 49 of these two external DMA channel units 54, 55 is connected to the processor bus 10.

35   The first external DMA channel unit 55 is connected via a link 47 to the first DTU 34. The

second external DMA channel unit 54 is connected via a link 46 to the second DTU 44. The internal channel 49 of these two external DMA channel units 54, 55 is connected to the processor bus 10.

The DTU 34 comprises a first processor interface 60 allowing a programming link 52 to be established via the processor bus 10 to the processor P1 (not shown in Figure 3). The DTU 34 further comprises a direct access unit core (DAU core) 62, and two external DMA channel interfaces 61 and 63. The external DMA channel interface 61 serves as interface to the external DMA channel unit 55 and the external DMA channel interface 63 serves as interface to the external DMA channel unit 57.

The DTU 44 comprises a first processor interface 50 allowing a programming link 51 to be established via the processor bus 20 to the processor P2 (not shown in Figure 3). The DTU 44 further comprises a direct access unit core (DAU core) 52, and two external DMA channel interfaces 51 and 53. The external DMA channel interface 51 serves as interface to the external DMA channel unit 56 and the external DMA channel interface 53 serves as interface to the external DMA channel unit 54.

The clock signal of the first processor P1 (clock1) is fed via a clock line 58 to the following units: external DMA channel unit 54, external DMA channel unit 55, external DMA channel interface 53, external DMA channel interface 61, and DAU core 62. The clock signal of the second processor P2 (clock2) is fed via a clock line 59 to the following units: external DMA channel unit 56, external DMA channel unit 57, external DMA channel interface 51, external DMA channel interface 63, and DAU core 52.

The processor P1 configures the first DTU 34 by means of the first processor interface 60. The DAU core 62 of the DTU 34 is the control logic for the two external channel interface units 61 and 63. The DAU core 62 furthermore performs the data transfers ideally enhanced by a first-in first-out (FIFO). The same way the processor P2 configures the second DTU 44 via the second processor interface 50. In both cases the external channels of the first DMA unit 45 use the resources of the internal DMA channel 49 on the processor bus 10, and the external channels of the second DMA unit 35 usethe resources of the internal DMA channel 39 on the processor bus 20.

As illustrated in Figure 3, the intercore communication system 29 provides for a clock decoupling. All the blocks are either clocked by the clock1 of the processor P1 or by the clock2 of the processor P2 such that the activity on one processor does not require simultaneous and equivalent activity on the other processor.

In cases where there is no phase and/or frequency relationship between the signals clock1 and clock2, the DAU cores 52, 62 can be implemented such that they are enabled

to provide safe data transfers by means of appropriate handshaking signals. These handshaking signals are active between the DAU core 52 and the external DMA channel interface 53 as well as

between the DAU core 62 and the external DMA channel interface 63.

5   The external DMA channel interfaces and/or the DAU cores can be standardised. In other words, each DTU or DMA, according to the present invention, may contain an identical functional core. Only the processor interface has to be adapted depending on the actual processor and/or processor bus employed. This leads to a reduced development time due to maximising of re-use and reduced verification effort.

10   According to the present invention, a DMA unit is connected via its internal interface to a processor bus and via its external interface to a DTU. The external interface may be 8 bits wide.

The processor interface has a programming input (e.g. input 52 in Figure 3), since this interface serves for the programming of the DTU in which it is comprised. The

15   processor interface does not require any data link to the processor bus, since any data exchanged is handled by the DTU's external DMA channel interfaces. The setup and configuration of the bi-directional channel is done by a processor by programming via the processor interface the respective DTU's DAU core.

The DTU 34, for instance makes use of the external DMA channel 47 in order to

20   transfer information (data and/or control information) to and from the shareable unit 13.

Another embodiment is illustrated in Figure 4. A system is illustrated that comprises a first processor P1, a first processor bus 70, and a first shareable unit 76 being connected to the first processor bus 70. There is a second processor P2, a second processor bus 80 and a second shareable unit 86 attached thereto. A first bi-directional communication channel

25   is establishable via a first DMA unit 83 with an external DMA channel 85. The first DMA unit 83 is connectable to the first processor bus 70. A first DTU unit 82 is provided. The first DTU unit 82 is connectable via the external DMA channel 85 to the first DMA unit 83. The DTU unit 82 is programmable by the first processor P1. The programming takes place via the processor bus 70 and a programming link 84. Furthermore, a master unit (master2) 81 is provided. This

30   master unit 81 serves as an interface between the first DTU 82 and the second processor bus 80. A second bi-directional communication channel is establishable via a second DMA unit 73 with an external DMA channel 75. The second DMA unit 73 is connectable to the second processor bus 80. A second DTU unit 72 is provided. The second DTU unit 72 is connectable via the external DMA channel 75 to the second DMA unit 73. The DTU unit 72 is programmable by the

35   second processor P2. The programming takes place via the processor bus 80 and a programming

link 74. Furthermore, a master unit (master1) 71 is provided. This master unit 71 serves as an interface between the second DTU 72 and the first processor bus 70. A master in the present context is a unit being able to initiate (and continue) data transfers on the processor buses. The masters therefore need to have access to some kind of arbitration (prioritization) on the buses

5   (this prioritization is not part of the present patent application). The masters have an addressing circuitry used to select the active device on the processor bus.

Another embodiment is illustrated in Figure 5. An intercore communication system 99 is provided that allows to establish two bi-directional channels between the two processor busses 90 and 100. The processor P1 may be a digital signal processor (DSP) core, and

10   the processor P2 may be a system controller (SC) core, for example. In the present embodiment, there is one common DTU unit 92 which for example comprises the functional elements of the DTU1 and DTU2 of Figures 2, 3, or 4. One part of this common DTU 92 is programmable by the processor P1, as indicated by the arrow 104, and the other part is programmable by the processor P2, as indicated by the arrow 94. Details of this DTU 92 are depicted in Figure 6.

15   The common DTU 92 comprises a first processor interface 120 allowing a programming link 104 to be established via the processor bus 90 to the processor P1 (not shown in Figure 6). The DTU 92 further comprises a direct access unit core (DAU core) 122, and two external DMA channel interfaces 121 and 123. The external DMA channel interface 121 serves as interface to the DMA1 unit 101 and the external DMA channel interface 123 serves as

20   interface to the DMA2 unit 93. The DTU 92 further comprises a second processor interface 110 allowing a programming link 94 to be established via the processor bus 100 to the processor P2 (not shown in Figure 6). The DTU 92 further comprises another direct access unit core (DAU core) 112, and two external DMA channel interfaces 111 and 113. The external DMA channel interface 111 serves as interface to the DMA2 unit 93 and the external DMA channel interface

25   113 serves as interface to the DMA1 unit 101. How the two clock signals clock1 and clock2 are applied is also shown in Figure 6.

The DTU 92 programming is preferably done using two separate register sets, each register set being assigned by one processor. P1 or P2. This allows to avoid conflicts with simultaneous accesses performed by the two DAU cores 112 and 122. However, a prioritisation

30   scheme is required that allows to prioritise requests from the processor P1 or requests from the processor P2. The following two schemes are proposed:

-       No priority is specified and the operation is based on the first come first serve principle, i.e., the processor that comes fist has the priority over the other processor. Ongoing transfers are always completed and new transfers are put on a waiting queue..

35   -   The processor P1 has priority over the processor P2, or vice versa. An ongoing

transfer of low priority data can be interrupted by a request submitted by the processor that has the higher priority. The interrupt of the transfer happens transparent to the low priority core. After the high priority request has finished, the low priority request is resumed. A high priority transfer is never interrupted.

5   According to the present invention, the DTU units make use of external DMA channels to transfer data to/from the shareable unit that is connectable to the processor bus of the other processor. Such an external DMA channel, contrary to the internal DMA channels which are programmed by the respective processor, are set-up by external agents in order to get access to the resources of the other processor. The external agents in this patent application are the

10   commands programmed by a remote processor to have access to a resource on the local processor - the internal DMA channels are programmed by the local processor itself.

            The present invention can also be employed in systems with more than two processors. A third processor might be connected via its own processor bus, a third DMA3 unit and a third DTU3 to the DMA2 unit of the second processor, for example. This would allow the

15   third processor to establish a bi-directional channel to resources that are associated with the second processor.

            In yet another embodiment of the invention, two or more processors and a communication channel for inter-processor communication in accordance with the present invention, are integrated into a custom application specific integrated circuit (ASIC).

20   It is an advantage of the architecture presented and claimed herein that it supports multiple heterogeneous processors. The inventive scheme can be expanded to suit general multi-core communication needs. Due to the present invention, the number of bus masters for each processor can be reduced, as potentially available DMA units can be used for this purpose. The concept and design reuse is another advantage. Different other advantages

25   have been mentioned in connection with the various embodiments of the present invention.

            The proposed architecture is symmetric and applicable to most microprocessor architectures. It can be expanded to multi-core architectures, i.e., it is independent of the number of cores.

            The present invention is well suited for use in computing devices, such as PDAs,

30   handheld computers, palm top computers, and so forth. It is also suited for being used in cellular phones (e.g., GSM phones), cordless phones (e.g., DECT phones), and so forth. The architecture proposed herein can be used in chips or chip sets for the above devices or chips for Blue tooth applications.

            It is appreciated that various features of the invention which are, for clarity,

35   described in the context of separate embodiments may also be provided in combination in a

single embodiment. Conversely, various features of the invention which are, for brevity, described in the context of a single embodiment may also be provided separately or in any suitable sub combination.

In the drawings and specification there has been set forth preferred embodiments
5 of the invention and, although specific terms are used, the description thus given uses terminology in a generic and descriptive sense only and not for purposes of limitation.